

Validité Informativ et Intentionalité

Joseph Sobel Ndiaye

Fondateur de

QUANTICS F <quantics-f.com>

Recherche & Développement en Sciences Cognitives

Mars 2020

Nous présentons les résultats de travaux d'application effectués dans le cadre de nos recherches en sciences cognitives.

*La problématique que nous abordons pourrait être résumée ainsi : **Comment extraire l'intention induite par un corpus de données qualitatives ?***

Nos résultats peuvent être exploités dans des domaines connexes aux sciences de l'information ; particulièrement la psycholinguistique et la théorie de la décision.

Nous analysons tout processus de cognition lexicale comme diffusion de corpus lexicaux dans un champ informatif.

- Par **corpus lexical**, nous entendons un ensemble formé par des unités lexicales couvrant une aire de signification.
- Par **champs informatif**, nous entendons un ensemble composé d'aires de signification.

La diffusion de corpus lexicaux dans un champ informatif a pour incidence une réalisation de signification.

Un processus de cognition lexicale augmente continuellement la stabilité (équilibre) de la relation **Signifiant**><**Signifié**.

Cette tendance à l'équilibre du lien **signifiant**><**signifié** est le caractère fondamental de toute **intentionalité** humaine.

+++^[P]_{SEP}

L'intention linguistique a donc une propriété asymptotique vers la Signification (information complète).

Cette **complétude de l'information** constitue un point d'équilibre théorique car en réalité, elle n'est pas atteinte du fait de la non-identité des humains.

Néanmoins, elle reste une tendance naturelle.

Nous avons exploité des corpus lexicaux composés de mots du langage naturel.

La tendance naturelle à la réalisation de signification (complétude de l'information), traduit -chez l'humain- une **aversion à l'incertitude**.

L'aversion à l'incertitude diffère de l'incertitude en ceci :

- L'incertitude est une éventualité non-mesurable.
- L'aversion à l'incertitude est mesurable selon le cas d'étude.

Dans le cas qui nous intéresse, à savoir le Langage naturel, notre approche consiste à utiliser une mesure de l'entropie de la structure sémantique de chaque mot composant un corpus lexical.

Cette mesure de **l'entropie cognitive** nous renseigne sur la caractéristique « **Signal/Bruit** » d'un corpus lexical.

La **validité informative** représente le potentiel d'un corpus lexical à "réaliser" une signification.

Cette "réalisation" de signification -ou *communication*- est incomplète car, comme mentionné plus-haut, la complétude de l'information ne peut être atteinte entre humains.

L'intentionnalité d'un champ informatif est donc fonction de la validité informative des corpus lexicaux qui la composent.

La connaissance de la validité informative des corpus lexicaux est un indicateur fiable de la **significativité** d'un champ informatif en langage naturel.

Nous pouvons donc obtenir une bonne connaissance de **l'intention implicite** de tout champ informatif en langage naturel.

ILLUSTRATION

Nous posons l'hypothèse suivante :

- *La langue française est une ontologie universelle en langage naturel. Il est possible de décrire n'importe quel concept grâce à des mots du langage naturel, pour qu'il soit compris par un maximum d'individus pratiquant cette même langue.*

> A l'échelle du mot, il est possible de spécifier un mot caché grâce à la connaissance graduée de ses synonymes.

De plus, la reconstitution optimale du mot caché est possible par une connaissance de ses synonymes selon une ordinalité liée à la validité informative de chaque synonyme.

Conséquence : Il existe au moins un synonyme du mot caché, dont la connaissance rend ce dernier plus facile à appréhender.

Exemple :

- le mot caché est «**éclair**»
- Ses synonymes sont : *éclat, étincelle, flamme, flash, foudre, fulguration, illumination, lueur, lumière, pâtisserie, révélation, tonnerre, trait.*
- Le calcul de validité sémantique des synonymes du mot «éclair», nous donne les valeurs ci-contre :

trait	0,0083
révélation	0,0512
lumière	0,0147
illumination	0,0582
foudre	0,0418
flamme	0,0086
éclat	0,0078
étincelle	0,0300
flash	0,3333
fulguration	0,0628
lueur	0,0328
pâtisserie	0,0340
tonnerre	0,0436

- Nous observons que le mot **flash** (0,3333) a une validité informative supérieure.

> A l'échelle de la phrase, il est possible de spécifier l'intention implicite (ou sujet caché).

Exemple :

- Phrase A : «**Le ciel est bleu.**»

ciel	0,5000
bleu	0,0154

- Phrase B : «**Science sans conscience n'est que ruine de l'âme.**»

science	0,0271
conscience	0,0209
ruine	0,0063
âme	0,0101

- Phrase C : «**Être philosophe ne consiste pas à savoir beaucoup de choses, mais à être tempérant.**»

être	0,0167
philosophe	0,0382
consiste	0,0883
savoir	0,0156
beaucoup	0,0067
chose	0,0252
tempérant	0,0676

CONCLUSION

Il existe pour toute intentionalité exprimée en langage naturel, des mots ayant une validité informative supérieure.

Ainsi, nous pouvons expliciter l'intention implicite de tout champs informatif en langage naturel.